Proc Mixed for Repeated Measures Data
(by a Non-statistician)

Jaswant Singh
Veterinary Biomedical Sciences
Most researchers use statistics the way a drunkard uses a lamp-post – more for support than illumination

- Winfred Castle
Treatment 1

Animal ID is nested within treatment

Our model statement for split-plot design will look like:

Model  \[ \text{Diameter} = \text{trt} \  \text{animal(trt)} \  \text{day} \  \text{trt*day} \]

Animals are considered a ‘random factor’ in the analysis

A random factor contains only a sample of the possible levels of the factor, and the intent is to generalize to all other levels
Student ID is nested within Method

Students are considered a ‘random factor’ in the analysis
Our questions are:

- How does the follicle diameter increase or decrease over time?
- How does one type of follicle differ from another in its response over time?
- Are there any interactions among follicles due to the treatment effect?
What do we want to accomplish?

- To analyze time series data
- To find out if there is a treatment effect
**Experimental designs**

**Experimental Units:** Animals

**Experimental Design:** Completely random or one-way

**Treatment Design:** Gradient

**Response Design:** Repeated

<table>
<thead>
<tr>
<th>Animal ID in each group</th>
<th>Control</th>
<th>Estradiol 0.1 mg i/m</th>
<th>Estradiol 0.5 mg i/m</th>
<th>Estradiol 1.0 mg i/m</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>12</td>
<td>69</td>
<td>11</td>
<td>68</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>71</td>
<td>23</td>
<td>70</td>
</tr>
<tr>
<td>21</td>
<td>21</td>
<td>9</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>34</td>
<td>34</td>
<td>6</td>
<td>33</td>
<td>5</td>
</tr>
<tr>
<td>58</td>
<td>58</td>
<td>31</td>
<td>57</td>
<td>30</td>
</tr>
</tbody>
</table>
SAS® PROC MIXED

- A new analysis tool which is appropriate for analyzing repeated measures data because it models the covariance of the data as well as the mean and the variance.
  - In repeated measures data, the data collected at one point in time is often not independent of the data collected at another time in the study (i.e., heterogeneity of residuals, the existence of covariance in your data set).

- Capable of analyzing data with missing values.
Correlation Matrices

<table>
<thead>
<tr>
<th>Independent</th>
<th>Compound Symmetry</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0 0 0</td>
<td>1 α α α</td>
</tr>
<tr>
<td>0 1 0 0</td>
<td>α 1 α α</td>
</tr>
<tr>
<td>0 0 1 0</td>
<td>α α 1 α</td>
</tr>
<tr>
<td>0 0 0 1</td>
<td>α α α 1</td>
</tr>
</tbody>
</table>

Independent: No correlation

Compound Symmetry: Fixed and constant Correlation within a subject is presumed

<table>
<thead>
<tr>
<th>Unstructured</th>
<th>Autoregressive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 α_{21} α_{31} α_{41}</td>
<td>1 α α^2 α^3</td>
</tr>
<tr>
<td>α_{21} 1 α_{32} α_{42}</td>
<td>α 1 α α^2</td>
</tr>
<tr>
<td>α_{31} α_{32} 1 α_{43}</td>
<td>α^2 α 1 α</td>
</tr>
<tr>
<td>α_{41} α_{42} α_{43} 1</td>
<td>α^3 α^2 α</td>
</tr>
</tbody>
</table>

Unstructured: Correlation matrix is completely unspecified

Autoregressive: Correlation decrease with distance
**Assumptions:**

- Data is normally distributed
- Variances and covariances of the data exhibit a structure matching one of those available in PROC MIXED
- Means (expected values) of the data are linear in terms of a certain set of parameters
Run PROC MIXED using different covariance structures:

- Compound Symmetry (CS)
- Huynh-Feldt (HF)
- Unstructured (UN, UN(1))
- Autoregressive (AR(1))

Select the model with the best fit

- The highest (i.e., most positive) values of Akaike’s Information and Schwarz’s Bayesian Criterion
First thing first

Test of Homogeneity of Variance

proc glm;
  class trt;
  model folliclediameter = trt;
  means trt / hovtest;
run;
The Syntax

data;
  input trt id day folliclediameter;
cards;
  . . . . . .
proc mixed;
  class id trt day;
  model folliclediameter = trt day trt*day / htype=3;
  repeated day / subject=id(trt) type=*;
run;

Run multiple analyses by replacing * with
  CS for Compound Symmetry
  HF for Huynh-Feldt
  UN, UN(1) for Unstructured
  AR(1) for Autoregressive
**SAS® PROC MIXED:**
Which model do we use?

<table>
<thead>
<tr>
<th></th>
<th>Akaike’s Criterion</th>
<th>Schwarz’s Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>-280.9</td>
<td>-283.2</td>
</tr>
<tr>
<td>AR(1)</td>
<td><strong>-245.1</strong></td>
<td><strong>-245.4</strong></td>
</tr>
<tr>
<td>UN</td>
<td>No fit</td>
<td>No fit</td>
</tr>
<tr>
<td>UN(1)</td>
<td>-306.2</td>
<td>-308.7</td>
</tr>
<tr>
<td>HF</td>
<td>No fit</td>
<td>No fit</td>
</tr>
</tbody>
</table>
To determine where differences exist:

- Least Squares Means
- Contrasts
- Estimates

In contrast to PROC GLM, PROC MIXED averages across the repeated measures and computes standard errors accounting for the appropriate covariance structure.
PROC GLM provides more extensive results for the traditional univariate and multivariate approaches to repeated measures.

PROC MIXED offers a richer class of both mean and variance-covariance models, and you can apply these to more general data structures and obtain more general inferences on the fixed effects.
References:

